



# ASSESSING PHYLOGENETIC SIGNAL WITH MEASUREMENT ERROR: A COMPARISON OF MANTEL TESTS, BLOMBERG ET AL.'S *K*, AND PHYLOGENETIC DISTOGRAMS

Olivier J. Hardy<sup>1,2</sup> and Sandrine Pavoine<sup>3,4</sup>

<sup>1</sup>*Evolutionary Biology and Ecology Unit, Faculté des Sciences, Université Libre de Bruxelles CP160/12, 50 av. F. Roosevelt, 1050 Brussels, Belgium*

<sup>2</sup>*E-mail: ohardy@ulb.ac.be*

<sup>3</sup>*Muséum national d'Histoire naturelle, Département Ecologie et Gestion de la Biodiversité, UMR 7204 CNRS UPMC, 61 rue Buffon, 75005 Paris, France*

<sup>4</sup>*Mathematical Ecology Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom*

Received July 8, 2011

Accepted February 22, 2012

In macroevolutionary studies, different approaches are commonly used to measure phylogenetic signal—the tendency of related taxa to resemble one another—including the *K* statistic and the Mantel test. The latter was recently criticized for lacking statistical power. Using new simulations, we show that the power of the Mantel test depends on the metrics used to define trait distances and phylogenetic distances between species. Increasing power is obtained by lowering variance and increasing negative skewness in interspecific distances, as obtained using Euclidean trait distances and the complement of Abouheif proximity as a phylogenetic distance. We show realistic situations involving “measurement error” due to intraspecific variability where the Mantel test is more powerful to detect a phylogenetic signal than a permutation test based on the *K* statistic. We highlight limitations of the *K*-statistic (univariate measure) and show that its application should take into account measurement errors using repeated measures per species to avoid estimation bias. Finally, we argue that phylogenetic distograms representing Euclidean trait distance as a function of the square root of patristic distance provide an insightful representation of the phylogenetic signal that can be used to assess both the impact of measurement error and the departure from a Brownian evolution model.

**KEY WORDS:** Abouheif proximity, Euclidean distance, intraspecific variability, Mantel test, measurement error, phylogenetic distogram, phylogenetic signal.

Characterizing phylogenetic signal, that is, the tendency of related species to resemble one another, is often a key point to assess macroevolutionary patterns in comparative biology (Losos 2008; Revell et al. 2008). To this end, many different approaches have been developed (e.g., Abouheif 1999; Diniz-Filho 2001; Freckleton et al. 2002; Blomberg et al. 2003; Ives et al. 2007; Pavoine et al. 2008, 2010; Zheng et al. 2009; Fritz and Purvis 2010; Ives

and Garland 2010; Jombart et al. 2010). They include the Mantel matrix correlation between a matrix of pairwise phylogenetic distances and of pairwise phenotypic distances between species. In a recent article, Harmon and Glor (2010) criticized the use of the Mantel test (1) to test for phylogenetic signal because it lacks power and (2) to test for the evolutionary correlation between two characters because it is then liberal. Although we think the second

statement is well justified, here we argue that the first one is not. We show realistic situations where the Mantel test proves more powerful to detect phylogenetic signal than a permutation test using Blomberg et al.'s (2003)  $K$  statistic, the alternative approach supported by Harmon and Glor (2010), especially when measurement errors (MEs) cannot be accounted for in the  $K$  statistic.

## Testing Power of the Mantel Test and the $K$ Statistic

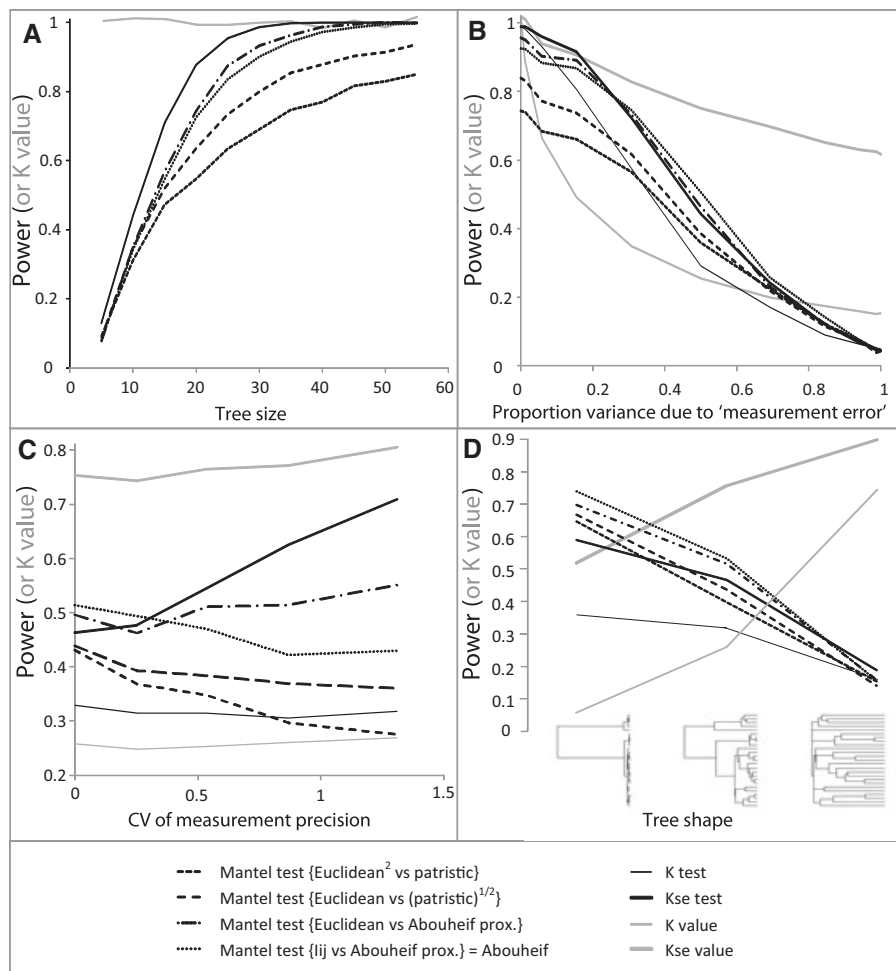
Testing for a phylogenetic signal using the Mantel test requires two matrices, one for trait distances and one for phylogenetic distances among species. Hereafter, we denote  $\mathbf{E}$  the matrix of Euclidean distances among species traits and  $\mathbf{P}$  the patristic distances among species on the phylogeny (sum of branch lengths linking species). Harmon and Glor (2010) considered, respectively,  $\mathbf{E}^2$  (squared Euclidean distances) and  $\mathbf{P}$  in their analysis of the power of the Mantel test because they are expected to vary proportionally under Brownian evolution. However, the choice of the distance measures affects the power of the Mantel test, a point neglected by Harmon and Glor (2010). To assess this point, we performed numerical simulations in the R statistical software (R Development Core Team 2011) and will mention hereafter the main functions and contributed R packages used to this end (see Supporting information for the R code).

As an alternative to  $\mathbf{E}^2$  for trait distances, one can consider  $\mathbf{E}$ , which is expected to increase proportionally with  $\mathbf{P}^{1/2}$  (the square root of patristic distance, computed using the function “cophenetic.phylo” from the R package “ape,” Paradis et al. 2004). Simulations generated under a birth–death process (using function “sim.bd.taxa” from the R package “TreeSim,” Stadler 2011) and where the evolution of a trait followed a Brownian motion model (using the function “rTraitCont” from the R package “ape,” Paradis et al. 2004) without ME (equivalent to Harmon and Glor 2010), show that the Mantel test (performed using the function “mantel” from the R package “vegan,” Oksanen et al. 2011) becomes more powerful when using  $\mathbf{E}$  and  $\mathbf{P}^{1/2}$  rather than  $\mathbf{E}^2$  and  $\mathbf{P}$  (Fig. 1A). Still higher testing power (Fig. 1A) can be achieved using an alternative phylogenetic distance derived from Abouheif (1999) and formalized by Pavoine et al. (2008), which can be computed using the function “proxTips,” method = “oriAbouheif,” from the R package “adephylo” (Jombart and Dray 2008) with a dependency on package “phylobase” (Hackathon et al. 2011). Abouheif proximity is  $1/2^n$  where  $n$  is the number of nodes separating species  $i$  and  $j$ . This formula is valid for a fully dichotomous tree, see Pavoine et al. (2008) for the case of polytomies. Nevertheless, under a Brownian evolution model (BM) and also under an Ornstein–Uhlenbeck model (OU; results not shown), the power of the Mantel test using Euclidean trait distances and the com-

plement of Abouheif proximity closely approaches but remains inferior to the power of the  $K$  statistic (Fig. 1A), which can be computed using the function “phylosig,” method = “K” and se = NULL, from the R package “phytools” (Revell 2011).

This pattern changes when traits display some level of “measurement error”, which can account both for limited measurement precision and actual phenotypic variation within species (Ives et al. 2007). To simulate this case, after evolving a trait under BM, replicates of trait measures were obtained for each species by adding normally distributed random terms to the intrinsic species trait value (model BM + ME). The  $K$  statistic was then estimated either using only the mean trait values per species and neglecting ME ( $K$ ) following Blomberg et al. (2003), or by integrating also the estimated standard errors (se) of the mean values for each species ( $Kse$ ) following Ives et al. (2007; using the R function “phylosig,” method = “K” and se = standard errors per species). As can be seen in Fig. 1B, for a dataset of 25 species where the variance of the error terms (i.e., the precision) is constant among species, the Mantel test using Euclidean distances and the complement of Abouheif proximity becomes more powerful than the test on  $K$  as soon as at least about 10% of the observed phenotypic variance is due to ME. When ME is accounted for, however, the test on  $Kse$  becomes nearly as powerful as the best Mantel tests. Taking into account the precision of phenotypic measures for each species in  $Kse$  becomes even more interesting when this precision varies among species because the test on  $Kse$  can then become much more powerful than any Mantel test (Fig. 1C). In our simulations with 25 species where the standard deviation (SD) of the error terms followed a log-normal distribution with a mean equal to the interspecific SD of the intrinsic phenotypic values, this occurred when the coefficient of variation of the log-normal distribution was above about 0.35 (Fig. 1C). The precision of the phenotypic measures of each species could also be taken into account in the Mantel test by redefining the Mantel correlation statistic and this might provide an increase in testing power similar to the one observed with  $Kse$ . However, this development is beyond the scope of the present article.

Besides the precision of phenotypic measures, another important factor affecting the power of all the tests is the shape of the phylogenetic tree. In particular, the power tends to decrease when most speciation events are concentrated near the root (Fig. 1D, right, obtained by transforming the original tree using the function “deltaTree” from the R package “geiger,” Harmon et al. 2009, with a parameter delta = 10). However, the relative impact varies among the tests and, when 50% of the observed phenotypic variance is due to ME, both the  $K$  and  $Kse$  tests can become less powerful than all Mantel tests when most speciation events are very recent (Fig. 1D, left, with delta = 0.1). The conclusion of Harmon and Glor (2010) about the low power of the Mantel test was thus based on an insufficient analysis of the different possible

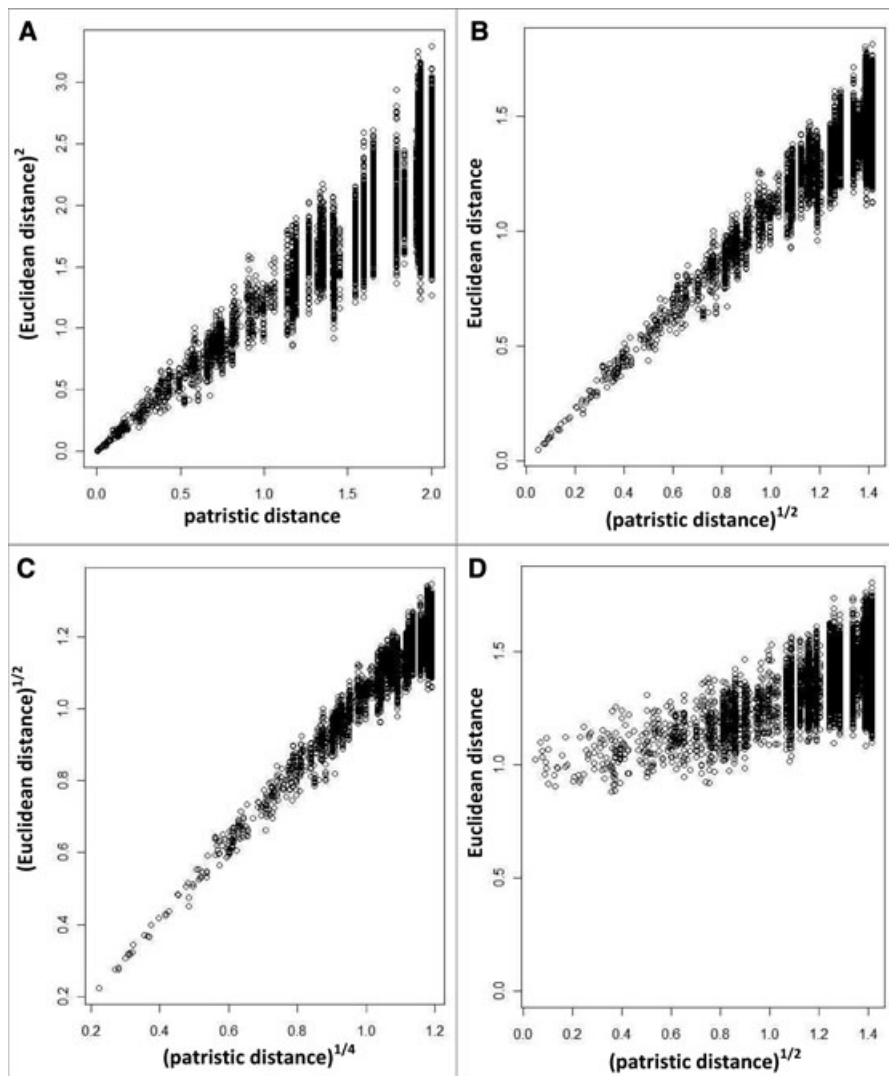


**Figure 1.** Statistical power ( $1 - \text{type II error}$ , for a nominal type I error rate  $\alpha = 0.05$ ) for tests of phylogenetic signal using the  $K$  and  $Kse$  statistics (thin and thick solid lines, respectively), and Mantel tests (discontinuous lines) with different trait and phylogenetic distance measures: squared Euclidean distance and patristic distance (as in Harmon and Glor 2010), Euclidean distances and square root (patristic distances), Euclidean distances and the complement of Abouheif proximity, the distance  $l_{ij}$  used in Moran's autocorrelation metric and Abouheif proximity (which gives the same power as Abouheif test). In each case, 1000 simulated datasets were analyzed for a range of tree sizes (A, where an original tree of 100 species was pruned to keep from five to 55 species) or for trees with 25 taxa (B, C, and D). Each tree was generated under a birth–death process model. The evolution of a single trait was simulated under Brownian motion model without measurement error (A; model BM) or with measurement error (B, C, and D; model BM + ME, simulating four measures per species) where the proportion of the phenotypic variance among species due to the error terms ranged from 0 to 1 (B) or was fixed to 0.5 (C and D). Although the precision (i.e., standard deviation of the error terms) of the phenotypic values were constant among species in (B) and (D), in (C) it varied according to a log-normal distribution with a coefficient of variation (CV) ranging from 0 to 1.3. In (D), the phylogenetic trees generated under a birth–death process (central) were transformed to concentrate speciation events late (right) or early (left) in the tree (with a parameter delta of 0.1 and 10, respectively; inserted trees represent examples for one replicate).

ways to apply the Mantel test and of the different scenarios of trait evolution.

The increased power of the Mantel test using  $E$  rather than  $E^2$  might result from a lowering of variance and/or an increase in negative skewness of interspecies trait distances (Fig. 2A, B). In fact, the power of the Mantel test can still be slightly increased by taking  $E^{1/2}$  against  $P^{1/4}$  under BM (Fig. 3). Under BM, the variance of the Euclidean distance increases with the patristic distance (Fig. 2B) so that the root transformations might improve the

power by stabilizing to some extent the variance (Fig. 2C). However, this improvement does not necessarily occur under BM + ME (Fig. 3) because the variance of the Euclidean distance is then less strongly related to patristic distance (Fig. 2D). It is worth noting that all Mantel tests considered, as well as  $K$  and  $Kse$  tests, displayed adequate type I error rate (i.e., rate of false positive) when trait values following a normal distribution were assigned randomly to each species (Fig. 1B when the proportion of variance due to ME reaches 1). Hence, the transformations of Euclidean

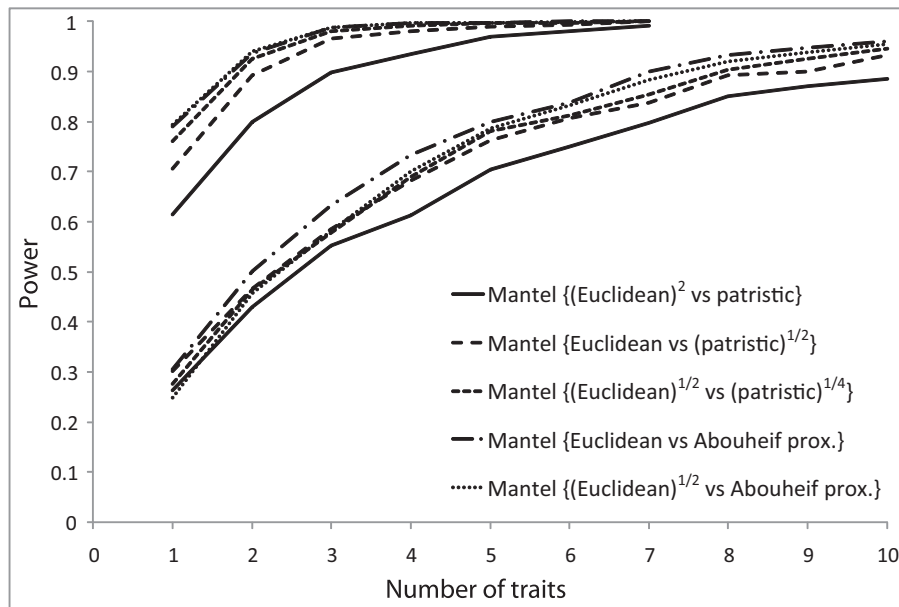


**Figure 2.** Relationship between Euclidean distance and patristic distance after different transformations for a tree of 200 taxa generated under a birth–death process and where 100 independent traits were simulated under a Brownian motion model without measurement error (A, B, C represent results from the same simulation run under model BM) or with measurement error accounting for 50% of the phenotypic variance for each trait (D, under model BM + ME). Note that the chosen transformations of Euclidean and patristic distances maintain linear relationships under the models simulated but they change the variance and the skewness of the distributions.

and patristic distances do not cause any trade-off between type I error rate and power (i.e.,  $1 - \text{type II error rate} = 1 - \text{rate of false negative}$ ).

The increased power brought by Abouheif proximity has already been shown by Pavoine et al. (2008) when comparing tests of phylogenetic signal based on Moran's  $I$  statistic (Abouheif proximity is here used in the weighting matrix of Moran's  $I$ ). In fact, the simulations conducted indicate that the Mantel test using Euclidean distances (E) and the complement of Abouheif proximities was slightly more powerful than the Moran's  $I$  test using Abouheif proximities under BM (Fig. 1A), but not necessarily under BM + ME (Fig. 1B). These differences should result from

the trait distance metric (in fact a trait proximity metric) implicit in Moran's  $I$  that is an autocorrelation coefficient defined as  $I_{ij} = (x_i - x_m)(x_j - x_m)/V(x)$ , where  $x_i$  and  $x_j$  are the trait values of species  $i$  and  $j$ , whereas  $x_m$  and  $V(x)$  are the mean and variance of the trait values over all species. Indeed, the same power is achieved using  $I_{ij}$  and Abouheif proximities in the Mantel test or Abouheif proximities in Moran's  $I$  test (i.e., Abouheif's test; Fig. 1A, B). In general, the highest power for the Mantel test was achieved using Abouheif proximities combined with Euclidean distance or  $I_{ij}$ . However, when most speciation events occur early in the tree (long branches to the tips) and traits evolve under BM, Mantel tests become more powerful using patristic distances than



**Figure 3.** Statistical power of Mantel tests with different trait and phylogenetic distance measures: (Euclidean distance)<sup>2</sup> and patristic distances, Euclidean distance and (patristic distances)<sup>1/2</sup>, (Euclidean distance)<sup>1/2</sup> and (patristic distances)<sup>1/4</sup>, Euclidean distances and the complement of Abouheif proximity, (Euclidean distance)<sup>1/2</sup> and the complement of Abouheif proximity. In each case, 1000 simulated datasets were analyzed for trees with 15 taxa generated under a birth–death process where the evolution of one to 10 independent traits was simulated under Brownian motion model without measurement error (the five curves above, model BM) or with 50% of variance per trait due to measurement error (the five curves below, model BM + ME).

Abouheif proximities (results not shown); although when ME is added, Mantel tests associated with Abouheif proximities have slightly higher power (Fig. 1D, right).

The fact that a phylogenetic distance neglecting branch lengths (Abouheif proximity depends only on tree topology) can confer more power than patristic distance along the phylogeny might seem counterintuitive. The Abouheif proximity between two species is inversely proportional to the number of equivalent tree configurations (obtained by switching the order of branches descending from each node) modifying the relative position of the two species. When testing for a phylogenetic signal by random permutations of the species among the tips of the phylogeny (as for Mantel and Moran's *I* tests), Abouheif proximity conveys information on the number of permutations that change the relative positions of the species for equivalent trees. In addition, the distribution of Abouheif proximity shows substantial skewness. These characteristics of Abouheif proximity may be involved in its good testing power, although a formal mathematical justification is still lacking.

### *Limitations of the *K* Statistic and the Importance of ME*

An important result brought by the simulations is that ME lowers substantially the power of Blomberg et al.'s (2003) *K*-test and affects this test more than the Mantel tests considered, eventu-

ally reversing their respective powers. *K*, which was designed to quantify deviation from BM (under which *K* = 1, which is indeed the estimated value in the absence of ME, Fig. 1A), is also substantially biased downward by ME (Fig. 1B) and this bias is stronger when speciation events are concentrated near the tips (Fig. 1D, left). This weakness of the original *K* statistic was addressed by Ives et al. (2007) who developed an estimator accounting for ME (*K*<sub>se</sub> statistic) and requiring several measures per species to estimate the standard error of the mean value of each species. Interestingly, accounting for ME in *K*<sub>se</sub> increases its power compared to *K*.

In our simulations, although *K* and *K*<sub>se</sub> were close to 1 under BM without ME (Fig. 1A), *K* dropped to about *K* = 0.6 with only one-tenth of variance due to ME and to about *K* = 0.25 with half of the variance due to ME (Fig. 1B), and in the latter case it even dropped to as low as *K* = 0.05 when most speciation events are recent (Fig. 1D, left). By contrast, *K*<sub>se</sub> remained much more stable but also displayed some downward bias, sometimes reaching *K*<sub>se</sub> = 0.5 (thick gray lines in Fig. 1B–D; see also Zheng et al. 2009). In fact, *K* is very sensitive to trait differences between very closely related species. Hence, *K* drops quickly when ME occurs or when the phylogeny is inaccurate among closely related species (as is often the case). However, ME and inaccurate phylogeny reconstruction can also counterbalance; this occurs when the branches lengths of the tips are overestimated because, under BM, this precisely mimics the effect of ME, so that *K* departs



less from unity. The downward bias of  $K_{se}$  under ME (Fig. 1B–D) originates from the fact that the standard errors of the mean trait values per species were estimated using four simulated measures per species and were thus not very precise. Indeed, in the simulated conditions of Fig. 1D (central tree) where half of the phenotypic variance is due to ME,  $K_{se} = 0.75, 0.90, 0.98$ , and  $1.00$  when the precision is estimated with 4, 10, 30, and 100 measures per species, respectively. Increasing the number of measures per species would not, however, reduce a bias in the  $K_{se}$  value due to inexact branch lengths in the phylogenetic tree.

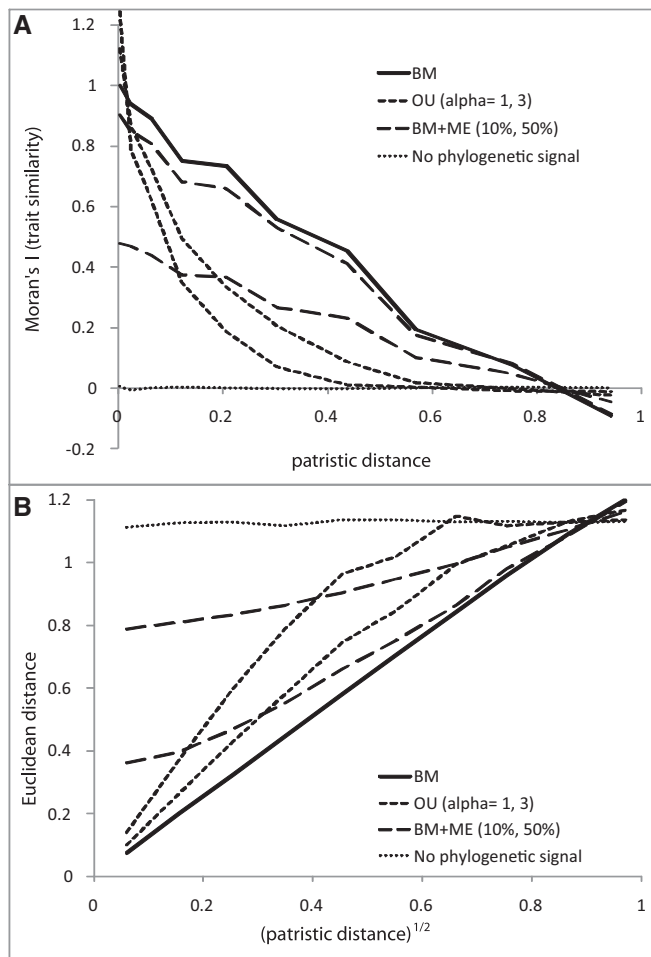
A limitation of  $K$  or  $K_{se}$  is that they are univariate while one may be interested in assessing the phylogenetic signal of a set of traits. By contrast, Euclidean distance can be estimated for a set of traits and the power to detect a signal using Mantel tests increases substantially with the number of traits (Fig. 3), a behavior also noticed by Zheng et al. (2009) who developed another multivariate test of phylogenetic signal.

Is ME important in real datasets? First, it must be noted that intraspecific variation is the rule for continuous quantitative phenotypic traits. Considering different measures of dispersal in European butterflies, Stevens et al. (2010) found that intraspecific variability represented from 11% to 133% of the interspecific variability. For adult body mass, a coefficient of variation of at least 0.1 or 0.2 among individuals of the same species is likely the rule for most animal species. If the tips of a phylogenetic tree represent genera or families rather than species, the associated variance would still increase further. When the interspecific variation of the sampled species is huge, as for the adult body mass of all mammals that range over about eight orders of magnitude (2 g for the bumblebee bat to 200 tons for the blue whale), ME can be negligible because after a logarithmic transformation of body mass to stabilize variances and approach normality, intraspecific phenotypic variation will remain very small in comparison to interspecific variation. However, when interspecific variation is limited, as for example among the body masses of wild canid species or for wood relative density that varies within about an order of magnitude (range = 0.11–1.39, mean  $\pm$  SD =  $0.65 \pm 0.18$  among about 2500 neotropical tree species, Chave et al. 2006), within-species variation can be comparatively substantial. ME can nevertheless be reduced by averaging measures taken over many individuals representative of the range of variation but this is not always feasible. For example, in Chave et al. (2006) study, wood densities of 60% of the species were based on a single reported measure and the average SD in species with multiple reported measures was 0.07, so that about 15% ( $0.07^2/0.18^2$ ) of the phenotypic variance would be due to ME if a single measure was available per species. This fraction would decrease proportionally to  $n^{-1/2}$  if  $n$  measures per species are available. This example shows a situation where a Mantel test applied using appropriate distance measures was shown to perform better than a test based

on the  $K$  statistic. Harmon and Glor (2010) might nevertheless be right in assuming that the Mantel test is not the most powerful way for detecting a phylogenetic signal but we need a more comprehensive study of the different tests proposed so far (see, e.g., Freckleton et al. 2002; Pavoine et al. 2008, 2010; Zheng et al. 2009; Fritz and Purvis 2010; Ives and Garland 2010).

## Alternative Ways to Characterize a Phylogenetic Signal: Phylogenetic Autocorrelograms and Distograms

When a phylogenetic signal is detected, researchers may be interested in inferring what process(es) generated the pattern. Different model-selection approaches have been developed to this end (Harmon et al. 2010). Blomberg et al.'s (2003)  $K$  is often used to assess whether there is less ( $K < 1$ ) or more ( $K > 1$ ) phylogenetic signal than under BM but it is important to realize that different phenomena can affect the  $K$ -value in the same way (e.g., both stabilizing selection in the OU process and ME reduce  $K$ ). Ives et al.'s (2007)  $K_{se}$  can at least partially resolve this by accounting for ME, provided that a standard error can be estimated for each species (e.g., several measures available). However, the actual process(es) that generated a particular pattern may be masked by standard model-selection approaches because a limited number of alternative models can be tested. Alternatively, phylogenetic autocorrelograms (average  $I_{ij}$  values over a set of phylogenetic distance intervals, Fig. 4A; Diniz-Filho 2001) or phylogenetic distograms (average Euclidean distance between species over a set of phylogenetic distance intervals, Fig. 4B) clearly show that contrasted patterns of phylogenetic signals are obtained when there is selection toward a unique optimum (OU model, Blomberg et al. 2003) or when ME occurs (BM + ME), differences that are not captured using the  $K$  statistic. For example, for the simulations presented in Figure 4,  $K \cong 0.34$  for the OU model with  $\alpha = 3$  as well as for the BM + ME model with 10% of phenotypic variance due to ME. Therefore, investigating patterns of trait evolution using phylogenetic autocorrelograms or distograms can be very informative and does not assume an explicit model. In the simulated examples, distograms using Euclidean distance against the square root of the patristic distance allowed distinguishing BM (linear relationship going through the origin), OU (concave relationship going through the origin) and BM + ME (linear relationship with a positive intercept) models (Fig. 4B). More generally, regressing pairwise  $I_{ij}$  or Euclidean distances between species on patristic or square root of patristic distances, respectively, can provide useful information on the nature of the phylogenetic signal. Indeed, the correlation coefficient, which can be tested using the Mantel test, would inform on the direction and strength of the signal, the intercept would



**Figure 4.** Phylogenetic autocorrelogram (A) and phylogenetic distogram (B) representing, respectively, the average similarity  $I_{ij}$  used in Moran's autocorrelation metric (A) or the average Euclidean distance (B) between taxa for a set of patristic distance intervals. Evolution of a single trait was simulated under different models on a phylogeny of 250 taxa (average values over 200 simulation replicates are given). BM: Brownian evolution model. OU: Ornstein–Uhlenbeck model with parameter  $\alpha = 1$  (curve closest to BM) or 3. BM + ME: Brownian evolution model including measurement error where the total phenotypic variance due to measurement error = 10% (curve closest to BM) or 50%. No phylogenetic signal: trait values randomly chosen from a normal distribution. Note that the scales of the x-axes are different for panels A and B but ensure in each case that the relationship is (near) linear under BM. The 10 patristic distance intervals were defined by their respective upper limits as  $(k/10)^2$  for interval  $k$  (1–10), after rescaling patristic distances so that the maximum value is 1.

quantify the degree of ME (in a broad sense including intraspecific variation), and the curvature of the relationship would quantify the degree of departure from BM after factoring out the effect of ME. Hence, distograms may illuminate some peculiarities of the processes yet to be explicitly modeled. They should push model

development in promising new directions and should reveal where current models are currently failing. This should be increasingly important as tree size used in comparative analysis is evermore increasing.

Interestingly, this approach also works using multiple traits because the Euclidean distance using all traits still increases proportionally to the square root of patristic distance (Fig. 2). It must be noted, however, that large phylogenies and/or many traits are required to limit the noise inherent in the stochasticity of evolutionary processes. Indeed, the relationships between pairwise Euclidean and patristic distances presented in Figure 2 are clear because 100 independent traits were simulated simultaneously but if a single trait was considered, the points would be dispersed in a triangular fashion because phylogenetically distant species could share similar trait values by chance. To clarify such noisy patterns, pairwise Euclidean distances can be averaged over a set of patristic distance intervals, as done in Figure 4B, but fairly large phylogenies are nevertheless needed to assess, for example, whether the relationship remains linear or not (departure from BM). Alternative approaches decomposing the phylogenetic signal among nodes have also been proposed by Jombart et al. (2010) and Pavoine et al. (2010).

## Conclusion

The power of the Mantel test depends on the metrics used to define trait distances and phylogenetic distances among species. Increasing power seems to be obtained by lowering variance and increasing negative skewness in phylogenetic and interspecific trait distances, for example, using Abouheif proximities combined with Euclidean distance or  $I_{ij}$ . In case of ME or phylogenetic inaccuracy among closely related species, Blomberg et al.'s (2003)  $K$  statistic is biased and the test of phylogenetic signal loses power but the approach of Ives et al. (2007) to account for ME largely corrects the bias and improves the power. This highlights the importance of having several phenotypic measures per species to be able to account for the precision in the mean phenotypic values. None of the tests compared was found to be superior in all simulated conditions with univariate data. However, the Mantel test can also be used with multivariate trait distances, increasing substantially its potential range of applications. We argue that phylogenetic distograms might constitute an insightful alternative for characterizing and exploring the phylogenetic signal without assuming a specific model of evolution, at least when dealing with large datasets, and may illuminate some peculiarities of the underlying evolutionary processes.

## ACKNOWLEDGMENTS

We thank K. Dexter, L. Harmon, and three anonymous reviewers for their constructive comments on previous drafts. OJH is a Research

Associate of the Belgian Fund for Scientific Research (F.R.S.-FNRS) which contributed to this project through grant F.4.519.10.F.

## LITERATURE CITED

- Abouheif, E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* 1:895–909.
- Blomberg, S. P., T. Garland Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Chave, J., H. C. Muller-Landau, T. R. Baker, T. A. Easdale, H. ter Steege, and C. O. Webb. 2006. Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecol. Appl.* 16:2356–2367.
- Diniz-Filho, J. A. F. 2001. Phylogenetic autocorrelation under distinct evolutionary processes. *Evolution* 55:1104–1109.
- Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* 160:712–726.
- Fritz, S. A., and A. Purvis. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* 24:1042–1051.
- Hackathon, R. B. B., M. Butler, P. Cowan, D. de Vienne, D. Edelbuettel, M. Holder, T. Jombart, S. Kembel, F. Michonneau, D. Orme, et al. (2011). *Phylobase*: base package for phylogenetic structures and comparative data. R package version 0.6.3. Available at <http://CRAN.R-project.org/package=phylobase> (accessed March 28, 2012).
- Harmon, L. J., and R. E. Glor. 2010. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* 64–67:2173–2178. doi: 10.1111/j.1558-5646.2010.00973.x
- Harmon, L. J., J. Weir, C. Brock, R. Glor, W. Challenger, and G. Hunt. 2009. *Geiger*: analysis of evolutionary diversification. R package version 1.3–1. Available at <http://CRAN.R-project.org/package=geiger> (accessed March 28, 2012).
- Harmon, L. J., J. B. Losos, J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. Kozak, M. A. McPeck, F. Moreno-Roark, T. J. Near, et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396. doi: 10.1111/j.1558-5646.2010.01025.x
- Ives, A. R., and T. Garland Jr. 2010. Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* 59:9–26.
- Ives, A. R., P. E. Midford, and T. Garland Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* 56:252–270.
- Jombart, T., S. Pavoine, S. Devillard, and D. Pontier. 2010. Putting phylogeny into the analysis of biological traits: a methodological approach. *J. Theor. Biol.* 264:693–701.
- Losos, J. B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* 11:995–1003.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2011. *Vegan*: community ecology package. R package version 2. 0–1. Available at <http://CRAN.R-project.org/package=vegan>. (accessed March 28, 2012).
- Paradis, E., J. Claude, and K. Strimmer. 2004. *APE*: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pavoine, S., S. Ollier, D. Pontier, and D. Chessel. 2008. Testing for phylogenetic signal in life history variable: Abouheif's test revisited. *Theor. Pop. Biol.* 73:79–91.
- Pavoine, S., M. Baguette, and M. B. Bonsall. 2010. Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecol. Monogr.* 80:485–507.
- R Development Core Team. 2011. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at <http://www.R-project.org> (accessed March 28, 2012).
- Revell, L. J. 2011. *phytools*: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* doi: 10.1111/j.2041-210X.2011.00169.x
- Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* 57:591–601.
- Stadler, T. 2011. *TreeSim*: simulating trees under the birth-death model. R package version 1.5. Available at <http://CRAN.R-project.org/package=TreeSim> (accessed March 28, 2012).
- Stevens, V. M., S. Pavoine, and M. Baguette. 2010. Variation within and between closely related species uncovers high intra-specific variability in dispersal. *PLoS ONE*, 5:e11123.
- Zheng, L., A. R. Ives, T. Garland Jr., B. R. Larget, Y. Yu, and K. Cao. 2009. New multivariate tests for phylogenetic signal and trait correlations applied to ecophysiological phenotypes of nine *Manglietia* species. *Funct. Ecol.* 23:1059–1069.

Associate Editor: J. Vamosi

## Supporting Information

The following supporting information is available for this article:

**Figure S1.** R code for the simulations presented in Figure 1.

**Figure S2.** R code for the simulations presented in Figures 2 and 4.

**Figure S3.** R code for the simulations presented in Figure 3.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.